PCT/AU98/00769



09/508713

**Patent Office**
**Canberra**

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PO 9221 for a patent by IMERSA PTY LIMITED filed on 16 September 1997.

I further certify that the above application is now proceeding in the name of LAKE DSP PTY LTD. pursuant to the provisions of Section 113 of the Patents Act 1990.

I further certify that the annexed specification is not, as yet, open to public inspection.

**PRIORITY DOCUMENT**
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

WITNESS my hand this Ninth
day of October 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

P/00/009
Regulation 3.2

AUSTRALIA

Patents Act 1990

# PROVISIONAL SPECIFICATION

Invention Title: Utilisation of Filtering Effects in Stereo Headphone Devices

The invention is described in the following statement:

GH REF: P25141-A/PJT:MB

**Utilisation of Filtering Effects in Stereo Headphone Devices**

Field of the Invention

The present invention relates to the creation of sound environments around a listener and, in particular, where the listener is listening to the sound environment via headphones.

Background of the Invention

A number of different sound reproduction techniques are in popular use. These techniques are created so as to provide a volumetric rendering of a sound such that it takes on spatial components. Historically, most sound was initially produced in a "mono" signal format. At present, however, one of the most popular formats is a stereo format wherein two sound signals are produced or transmitted such that, when output on a pair of speakers, they appear to have a spatial component or environment out of the front of a listener when those speakers are placed in front of the listener.

Unfortunately, when standard headphones are utilised, the out-of-head perception is lost and the sound appears to be coming from somewhere inside the listeners head and is substantially centralizsed.

Other sound formats face similar problems when reproduced over headphones. For example, the Dolby AC-3 format, another popular format, is designed for the placement of a number of speakers around a listener so as to create a substantially richer sound environment. Again, when headphone devices are utilised in such an environment the intended spatial location of the sound is lost and again the sound appears to come from within the head of a listener.

Summary of the Invention

It is an objection of the present invention to provide an improved method and system which allows for the playback of audio through headphones so as to create the illusion of sound sources external to the listener's cranium. The

system includes improvements which relate to the reduction in computational requirements of existing systems and improving the realism of the virtual speaker systems. The system provides for the production of a stable illusion of sound sources positioned around the user with an impression of a depth and distance and thereby provides a richer environment for the headphone listener.

In accordance with the first aspect of the present invention there is provided an apparatus for creating, utilizing a pair of oppositely opposed headphones, the sensation of a sound source being spatially distant from the area between the pair of headphones, the apparatus comprising a series of audio inputs representing audio signals being projected from an idealized speaker located at a spatial location relative to an idealized listener; a first mixing matrix means interconnected to the audio inputs for outputting a predetermined combination of the audio inputs as intermediate output signals; a filter for filtering the intermediate output signals and outputting filtered intermediate output signals; and a second mixing matrix means combining the filtered intermediate output signals to produce left and right channel stereo outputs.

Preferably, the first mixing matrix means outputs a linear combination of the audio inputs. The first matrix means can also apply a time varying gain to the audio inputs. The filters are preferably independent of one another. Approximations of Head Related Transfer Functions (HRTFs) are used which ordinarily would result in a limited system. However, the addition of a sparse tap reverb filter significantly improves the perception of the virtual sound direction and depth.

The present invention can be advantageously utilized in the processing of Dolby AC-3 inputs or stereo inputs, in addition to other standard formats.

In accordance with the second aspect of the present invention there is provided an audio processing method for

converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the inputs so as to create the appearance of sound located around a listener, the method comprising filtering each of the Dolby AC-3 inputs utilising first filters constructed to simulate the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear; applying a second filter to each of the inputs to simulate the reverberant tail of a suitably arranged virtual speaker to a corresponding listener's ear; and adding together the outputs from the filtering step and the applying step to produce left and right stereo headphone outputs.

Preferably, the first filters comprise short filter lengths whereas the second filters comprise substantially longer filter lengths. For example, the first filters can be about 2,000 taps in length and the second filters can be about 32,000 taps in length.

In accordance with the third aspect of the present invention there is provided an audio processing apparatus for converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the inputs so as to create the appearance of sound located around a listener, the apparatus comprising a first series of early response filters for filtering the inputs so as to produce outputs simulating the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear; a second series of reverberant tail filters for filtering the inputs so as to produce outputs simulating the reverberant tail response from a suitably arranged virtual speaker to a corresponding listener's ear; and a left and right output combining means for combining the outputs of the first and second series of filters so as to produce left and right headphone outputs. Preferably, the number of reverberant tail filters is two and the inputs are summed together before input to the

reverberant tail filters.

In accordance with the fourth aspect of the present invention there is provided a method of processing stereo input sound sources for playback over headphones so as to create the sensation of sound originating from around a headphone listener, the method comprising the steps of producing sum and difference signals from the stereo input sound sources; applying a direct ear response and shadow ear response filter to the difference signal to form a filtered difference output; applying a direct ear response, a shadow ear response and a reverberant response filter to the sum signal to form a filtered sum output; forming a first headphone output from the addition of the filtered difference output and the filtered sum output; and forming a second headphone output from the subtraction of the filtered difference output and the filtered sum output.

Preferably, the responses simulate head related transfer functions for the placement of virtual speakers at 30 degrees to the horizontal plane. The shadow ear response filter can comprise a short FIR filter matching the frequency response and group delay of a signal derived from deconvolving a direct ear response from an appropriate shadowed response.

The preferred embodiments include techniques which consist of input mixing, filters and output mixing. The present invention includes unique combinations of mixing techniques and filter techniques, including filter techniques which improve pscho-acoustic perceptions of spatial sound.

Brief Description of the Drawings

Notwithstanding any other forms which may fall within the scope of the present invention, preferred forms of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Fig. 1 illustrates the operation of a system of the

present invention;

Fig. 2 illustrates a generalised form of the preferred embodiment;

Fig. 3 illustrates a more detailed schematic form of the preferred embodiment;

Fig. 4 illustrates a schematic diagram of a Dolby AC-3 to stereo headphone converter;

Fig. 5 illustrates a stereo input to stereo output embodiment in schematic form; and

Fig. 6 illustrates in schematic form, one form of conversion from Dolby AC-3 inputs to stereo outputs in accordance with the present invention.

Description of Preferred and Other Embodiments

A number of the embodiments of the present invention will be described for different sound formats.

Turning initially to Fig. 1, there is provided a schematic illustration of the operation of an embodiment of the invention. In this embodiment, a series of audio inputs 11 are provided to a mechanism 12 which would normally form part of the prior art taking the audio signal inputs and creating a series of speaker feeds 13. The speaker feeds 13 can be provided for the various output formats, for example stereo output formats or AC-3 output formats. The operation of the portion within dotted line 14 being entirely conventional. The speaker feeds are forwarded to the headphone processing system 15 which outputs to a set of standard headphones 16 so as to simulate the presence of a number of speakers around the listener using headphones 16.

Fig. 1 illustrates the example where headphone processing system 16 simulates the presence of two virtual speakers 17, 18 in front of the user of headphones 16 as would be the normal stereo response. The arrangement of Fig. 1 has particular advantages in that it can be incorporated in any system that is generally utilised for the playback of stereo audio. The system processes the usual signals intended for playback over speakers and is

therefore compatible with and can be used in conjunction with any other system designed for enhancing the reproduction of audio over loudspeakers.

The general structure of a first form and implementation can be represented by a filter structure where each of the intended speaker feeds is passed through two filters, one for each ear. The resultant sum of all these filters is the signal sent to the appropriate headphone channel for that ear. In alternative embodiments, the filters may or may not be updated to reflect changes in the orientation of the listener's head inside the virtual speaker array. By updating the filters based on the physical orientation of a listener's head, an imersive head-tracked environment can be created. Various implementations can be variations on this theme so as to reduce computational requirements. Further, non-linear, active or adaptive components can be added to the structure to improve performance.

An example of the general structure in a more complex form is illustrated in Fig. 2. The implementation 20 includes a series of speaker feeds e.g. 21 each of which has a separate filter e.g. 22, 23 applied with one filter 22 being applied for a left hand channel and one filter 23 being applied for a right hand channel. The filter outputs are summed e.g. 24 together to form a final output 25.

The arrangement of Fig. 2 can lead to overburdening complexity in a large number of filters e.g. 22 must be provided which is likely to substantially increase costs. A technique for significantly reducing the computational requirements by taking advantage of symmetry is to utilise "shuffling" techniques. For a pair of channels, this represents applying filters to the sum and difference of the channels before recombination. For the stereo case where the filters are symmetric (i.e. FilterLL = FilterRR, FilterLR = FilterRL) this can reduce the computational requirements by 50%. This technique can be represented by

inserting a linear matrix mix before and after the filter banks.

More generally, as indicated in Fig. 3, the implementation structure 30 consists of:

- A number of inputs 31
- A mixing matrix 32 to produce a set of signals each of which is a linear combination of the input signals (note the intermediate set of signals may include the input signals themselves and may include duplicate signals).

In alternative embodiments, the matrix gains may be time varying.

- A filter e.g. 33 on each of the intermediate signals. The filters can be independent and thus can have different structures, lengths and delays (for example IIR, FIR, sparse tap IR, and low latency convolution).
- A mixing matrix 35 to combine the filtered intermediate signals appropriately to create the two headphone output signals 36.

Some specific implementations of the above general system are as follows:

High End AC-3 Decoder

As illustrated in Fig. 4, the Dolby™ AC-3 standard defines a set of 5 (.1) channels to be used as speaker feeds 41. These channels are derived from an AC-3 bit stream data source using an AC-3 decoder. Once decoded, the speaker feeds are suitable for utilisation as inputs 41 to the arrangement 40 of Fig. 4 which produces headphone outputs 42. Each of the five speaker feeds is passed through a filter e.g. 43, 44 for either ear and summed e.g. 45 to produce the headphone signal - making a total of 10 filters.

To achieve a high level of quality in the simulation of a virtual speaker array, fairly long filters are required to take into account the spatial geometry of the listening environment. With proper filter sets (incorporating equalisation for the headphones and proper head related

transfer functions) the results provide close to a perfect illusion of a set of external speakers being used.

The 10-filter design can be refined to reduce computational power without too much quality degradation by using 10 shorter filters and only two full-length filters. The two longer filters 47, 48 can be a binaural simulation of the tail of an average room response. A combination of all 5 speaker feeds is fed via summer 49 into the binaural tail filters 47, 48 to give an approximation of the real room response. Each of the short filters e.g. 43, 44 can be the early part of the response for that particular speaker to the listener's ear.

The filter length used in prototype implementations is typically 2000 taps at 48kHz sampling rate for the short filters e.g. 43, 44 and 32000 taps for the longer filters 47, 48. The long filters usually have a lower bandwidth and can be implemented with latency - this can be taken advantage of using a reduced sample rate processing to lower the computational requirements. The filters can be implemented using low latency convolution algorithms to lower the system latency and computational requirements.

The filter sets can be obtained by simulating a virtual speaker set-up using acoustic modelling packages such as CATT acoustics or by using a real or synthetic head placed inside a real speaker array.

The High End AC-3 decoder 40 provides a fairly accurate simulation through headphones of a virtual speaker array, however, it also includes a large amount of computational resource.

Low End Stereo Decoder

The Low-End Stereo Decoder as illustrated 50 in Fig. 5, is a device utilising only some of the features of the high-end computationally resourced system. The main aim is to manipulate a stereo source 51 for playback over headphones 52 to give the impression of the sound originating from around the listener, simulating the experience of listening

to a well configured stereo. The system is designed to be suitable for mass production at a low cost; thus the more important issues of the design are in reducing the computational complexity.

As noted previously, the general structure of the low-end stereo decoder 50 has two inputs 51 for conventional stereo and two outputs 52 for the headphone signals. A bank of two filters is used, operating on the sum 55 and difference 56 signals of the input stereo pair 51.

The low end stereo decoder 50 is another example, consistent with the general implementation outlines previously. In this case the matrix operations are a two channel sum 55 and difference 56 shuffle. The filters are applied to the sum and difference signals to half the computational requirements where the desired result is symmetric (i.e. L->L=R->R and L->R=R->L).

The performance of this system is dependent on the choice of filter coefficients. To reduce the computational requirements, short filters are ideally used. It has been found that the difference filter can be somewhat shorter than the sum filter and still produce a reasonable result.

The preferred form is to use a set of filters that is a combination of the head related transfer functions for 30° in the horizontal plane, and a semi-reverberant tail but fairly sparse filter. The filter construction can be as follows:

Given the following impulse responses

D Direct ear response - normalised to unity energy

S Shadowed ear response - scaled in proportion to D

R Reverberant response - normalised to unity energy

and the following parameter

$\alpha$ Presence - the amount of reverberant feed in the mix

then the following filters are applied to the sum and difference signals to produce new sum' and Diff' signals

$$Sum' = \left(\sqrt{\left(1-\alpha^2\right)}(D+S) + \alpha R\right) \otimes Sum$$

$$Diff' = \left(\sqrt{\left(1-\alpha^2\right)}(D-S)\right) \otimes Diff$$

To further reduce the amount of processing required a number of approximations can be made to the filter set. The direct ear response is assumed to be unity. The shadowed ear response can be approximated by a 5 tap FIR matching the frequency response and group delay of the exact signal derived from deconvolving a direct ear response from the appropriate shadowed response. Around 20 sparse taps can approximate the reverberant response from a 5-10ms delay line.

With this approach it has been found that the coefficients can be heavily quantised and reasonable performance maintained. The sum filter can be implemented as a set of 25 taps from a 256 tap delay line (at 48kHz) while the difference filter can be mere 6 taps from a 30 tap delay line. This allows the system to be implemented using around 3 MIPS thus making it suitable for low cost, mass production and incorporation into other audio products using headphones.

Further extensions to the implementation 50 can include:

- The use of low-latency convolution to allow the possibility of longer filters.
- The addition of further inputs and similar budget processing to allow for the simulation of "surround sound" formats. For example, a surround channel could be added that simulates the presence of sounds behind or around the rear of the listener.
- Incorporation of budget head tracking processing to change the early HRTF components to give a sense of stationary sound sources when the head is rotated.
- Addition of non-symmetric components to provide better

performance when the stereo signal has significant mono components in the mix.

- Addition of non-linear components to enhance the performance (for example a dynamic range compressor to improve the quality of listening in a noisy environment).

It can therefore be seen that the preferred embodiments utilise a unique combination of input mix-processing, filters and output mix-processing to create the appearance of 3-dimensional sound over headphones. The arrangements disclosed include reduced computational complexity and memory requirements resulting in a significant reduction in implementation costs. The filter structures and coefficients improve the directionality and depth of the sound with minimal increase in computational complexity. The simple HRTF approximations require little processing power having been significantly reduced from the normal 50-60 filter taps.

The significant HRTF features include

(a) the significant main energy component of the direct response (short time approximation) and the approximation of the convolution mapping of the direct response to the shadow or reflected response.

(b) the use of filter coefficients comprising a 5-10ms sparse tap filter after about 50-100 taps. The use of the reverberent filter enhances the performance of the HRTF approximations, normal HRTF's and room impulse responses by increasing the localisation and depth of sound.

(c) In a modification, the HRTF approximations can include coefficients for containing anti-phase component in the shadow response so as to improve rear localisation.

(d) The filters of preferred embodiments include a first part which provides directionality and

localisation and a second part which provides ambiance and room acoustics but minimal directionality.

The utilisation of the delivery format of the preferred embodiments provides considerable flexibility in the trade off of optimal computation and memory usage versus performance.

The extension of the system 50 of Fig. 5 to Dolby AC-3 inputs can be as shown 60 in Fig. 6. The center channel 61 is added 62, 63 to the front left and rear right channels respectively. The output signals are fed to delay units 64, 65 which can be 5 to 10 msec delay lines, before being fed to HRTFs 67 - 69 which provide outputs for summing 70, 71 to the left and right ears. The rear signals 73, 74 are used to form sum and difference signals 76,77 which are fed to HRTFs 79, 80 which provide anti-phase to the summing units 70, 71.

It would be further appreciated by a person skilled in the art that numerous variations and/or modifications any be made to the present invention as shown in the specific embodiment without departing from the spirit or scope of the invention as broadly described. The present embodiment is, therefore, to be considered in all respects to be illustrative and not restrictive.

We Claim:

1. An apparatus for creating, utilizing a pair of oppositely opposed headphones, the sensation of a sound source being spatially distant from the area between said pair of headphones, said apparatus comprising:

(a) a series of audio inputs representing audio signals being projected from an idealized speaker located at a spatial location relative to an idealized listener;

(b) a first mixing matrix means interconnected to said audio inputs for outputting a predetermined combination of said audio inputs as intermediate output signals;

(c) a filter system for filtering said intermediate output signals and outputting filtered intermediate output signals; said filter system including separate filters for filtering the direct response and short time response and an approximation to the reverberent response; and

(d) a second mixing matrix means combining said filtered intermediate output signals to produce left and right channel stereo outputs.

2. An apparatus as claimed in claim 1 wherein said first mixing matrix means outputs a linear combination of said audio inputs.

3. An apparatus as claimed in claim 1 wherein said first matrix means applies a time varying gain to said audio inputs.

4. An apparatus as claimed in any previous claim wherein said filters are independent of one another.

5. An apparatus as claimed in any previous claim wherein said audio inputs comprise Dolby AC-3 inputs.

6. An apparatus as claimed in any previous claim 1 to 4 wherein said audio inputs comprise stereo inputs.

7. An audio processing method for converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the

inputs so as to create the appearance of sound located around a listener, said method comprising:

filtering each of the Dolby AC-3 inputs utilising first filters constructed to simulate the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear;

applying a second filter to each of said inputs to simulate the reverberant tail of a suitably arranged virtual speaker to a corresponding listener's ear; and

adding together the outputs from said filtering step and said applying step to produce left and right stereo headphone outputs.

8. A method as claimed in claim 7 wherein said inputs are summed before being input to said second filters.

9. A method as claimed in claim 7 wherein said first filters comprise short filter lengths whereas said second filters comprise substantially longer filter lengths.

10. A method as claimed in claim 9 wherein said first filters are about 2,000 taps in length and said second filters are about 32,000 taps in length.

11. An audio processing apparatus for converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the inputs so as to create the appearance of sound located around a listener, said apparatus comprising:

a first series of early response filters for filtering said inputs so as to produce outputs simulating the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear;

a second series of reverberant tail filters for filtering said inputs so as to produce outputs simulating the reverberant tail response from a suitably arranged virtual speaker to a corresponding listener's ear; and

a left and right output combining means for combining the outputs of said first and second series of filters so as to produce left and right headphone outputs.

12. An audio processing apparatus as claimed in claim 11 wherein the number of reverberant tail filters is two and said inputs are summed together before input to said reverberant tail filters.

13. A method of processing stereo input sound sources for playback over headphones so as to create the sensation of sound originating from around a headphone listener, said method comprising the steps of:

(a) producing sum and difference signals from said stereo input sound sources;

(b) applying a direct ear response and shadow ear response filter to said difference signal to form a filtered difference output;

(c) applying a direct ear response, a shadow ear response and a reverberant response filter to said sum signal to form a filtered sum output;

(d) forming a first headphone output from the addition of said filtered difference output and said filtered sum output; and

(e) forming a second headphone output from the subtraction of said filtered difference output and said filtered sum output.

14. A method as claimed in claim 13 wherein said responses simulate head related transfer functions for the placement of virtual speakers at substantially 30 degrees to the horizontal plane.

15. A method as claimed in claim 13 wherein said filters comprise forming the following outputs:

$$Sum' = \left(\sqrt{\left(1-\alpha^2\right)}(D+S)+\alpha R\right) \otimes Sum$$

$$Diff' = \left(\sqrt{\left(1-\alpha^2\right)}(D-S)\right) \otimes Diff$$

where:

Sum and Diff are the sum signal and difference signal respectively;

Sum' and Diff' are the filtered sum output and filtered

difference output respectively;

D is the direct ear response - normalised to unity energy;

S is the shadowed ear response - scaled in proportion to D;

R is the reverberant response - normalised to unity energy;

$\alpha$ is the presence - the amount of reverberant feed in the mix.

16. A method as claimed in claim 13 wherein in said shadow ear response filter comprises a short FIR filter matching the frequency response and group delay of a signal derived from deconvolving a direct ear response from an appropriate shadowed response.

17. A method as claimed in claim 13 wherein said reverberant response filter approximates a delay line of between 5 - 10 ms

DOLBY AC3

18. A method of processing Dolby AC-3 input sound sources for playback over headphones so as to create the sensation of sound originating from around a headphone listener, said method comprising the steps of:

(a) producing sum and difference signals from the Right Rear and Left Rear input signals;

(b) producing an intermediate front left signal from the addition of the front left signal and the center right signal;

(c) producing an intermediate front right signal from the addition of the front right signal and the center signal;

(d) applying separate HRTF signals to said intermediate signals;

(e) applying an anti-phase HRTF to said sum and difference signals;

(f) summing the outputs of steps (d) and (e) to produce left and right channels headphone signals.

19. A method as claimed in claim 19 wherein said intermediate signals are delayed before the application of said HRTFs.

Dated this 5th day of September 1997

Lake DSP Pty Ltd
By their Patent Attorneys
GRIFFITH HACK

19. A method as claimed in claim 19 wherein said intermediate signals are delayed before the application of said HRTFs.

Dated this 5th day of September 1997

Lake DSP Pty Limited.
~~Imersa Pty. Ltd.~~

By their Patent Attorneys
GRIFFITH HACK

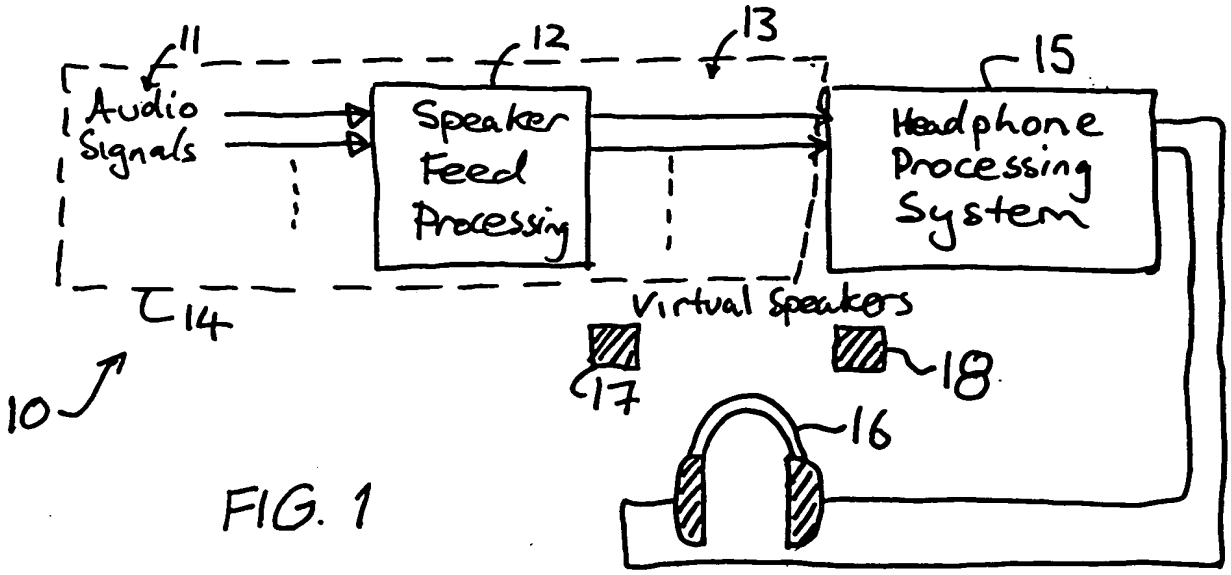
11
12
13
15
Audio Signals
Speaker Feed Processing
Headphone Processing System
14
Virtual Speakers
17
18
16
10

FIG. 1

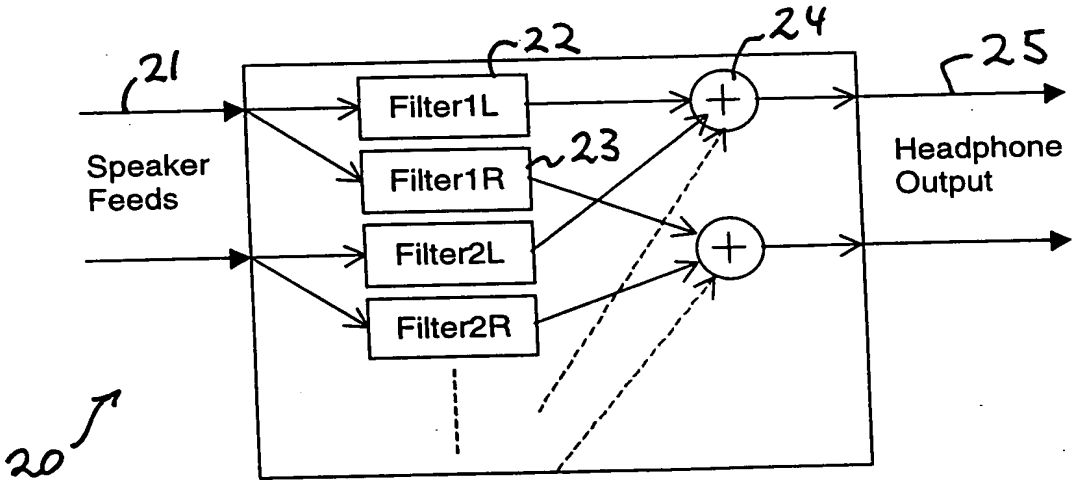
21
22
24
25
Filter1L
Filter1R
23
Filter2L
Filter2R
Speaker Feeds
Headphone Output
20

FIG. 2

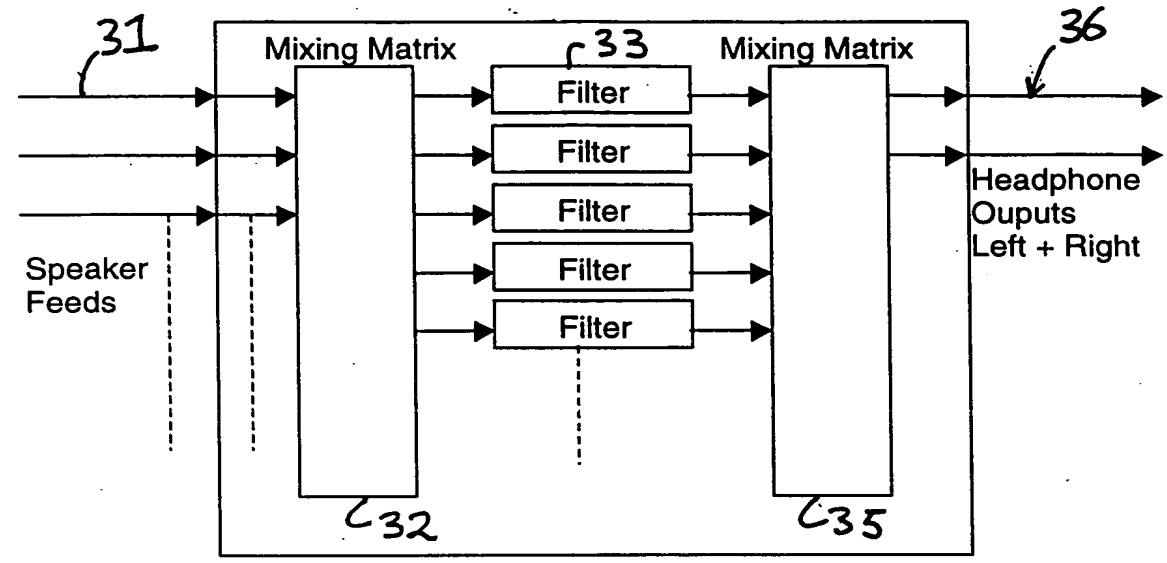
31
Mixing Matrix
33
Mixing Matrix
36
Filter
Filter
Filter
Filter
Filter
Speaker
Feeds
Headphone
Ouputs
Left + Right
32
35
30

FIG. 3

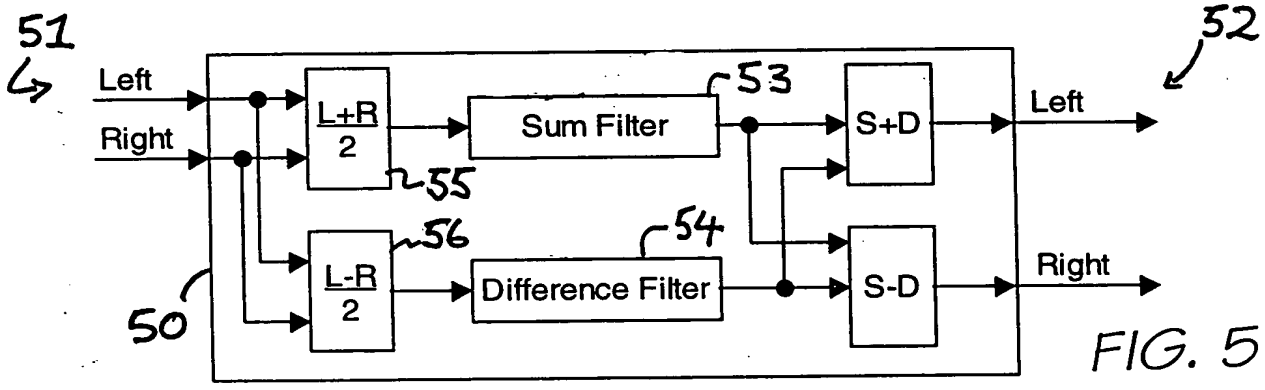
51
52
Left
Right
L+R
2
55
Sum Filter
53
S+D
Left
56
54
L-R
2
Difference Filter
S-D
Right
50

FIG. 5

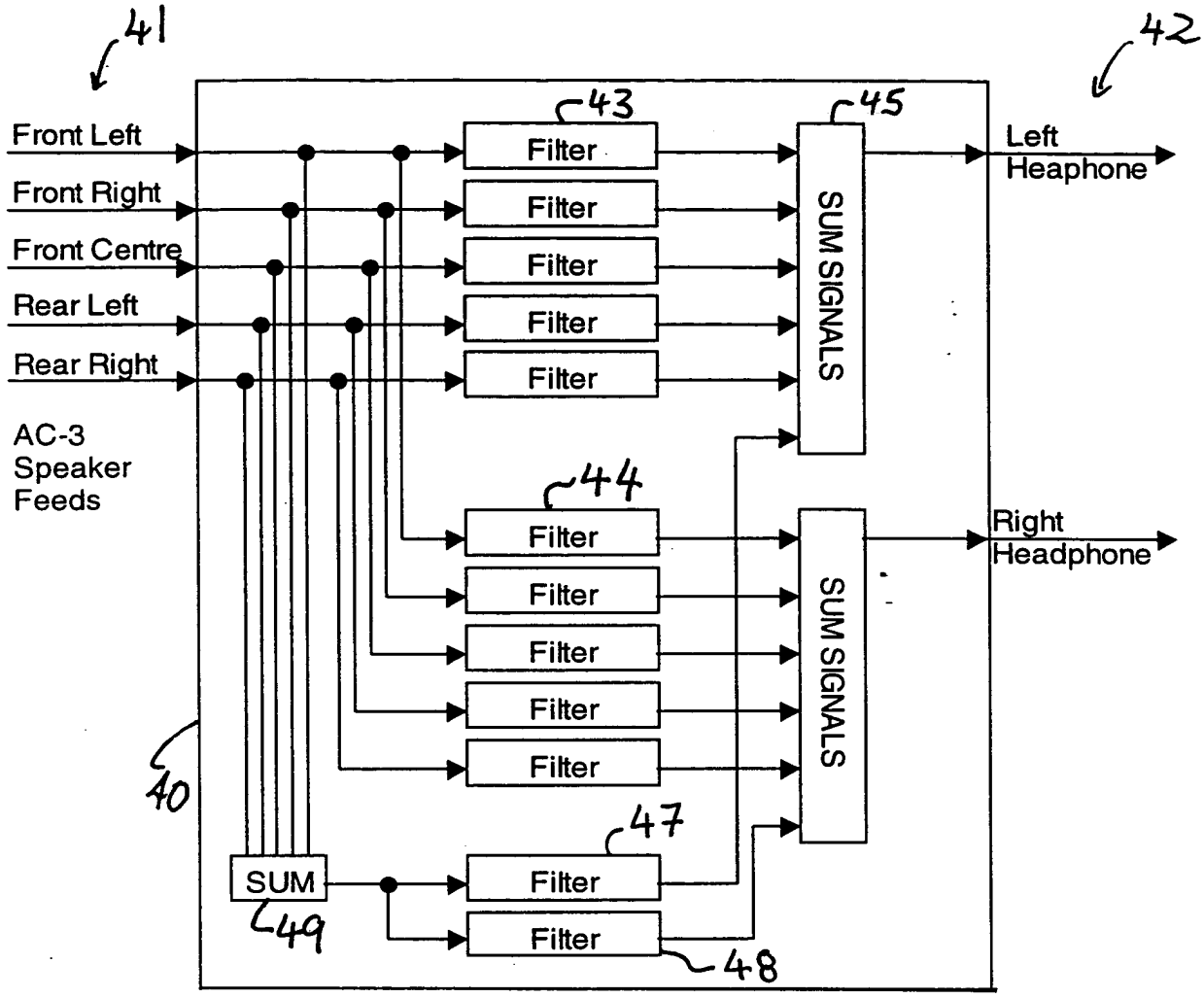
41
42
Front Left
Front Right
Front Centre
Rear Left
Rear Right
AC-3
Speaker
Feeds
43
Filter
Filter
Filter
Filter
Filter
45
SUM SIGNALS
Left
Heaphone
44
Filter
Filter
Filter
Filter
Filter
SUM SIGNALS
Right
Headphone
40
47
SUM
Filter
Filter
49
48

FIG. 4

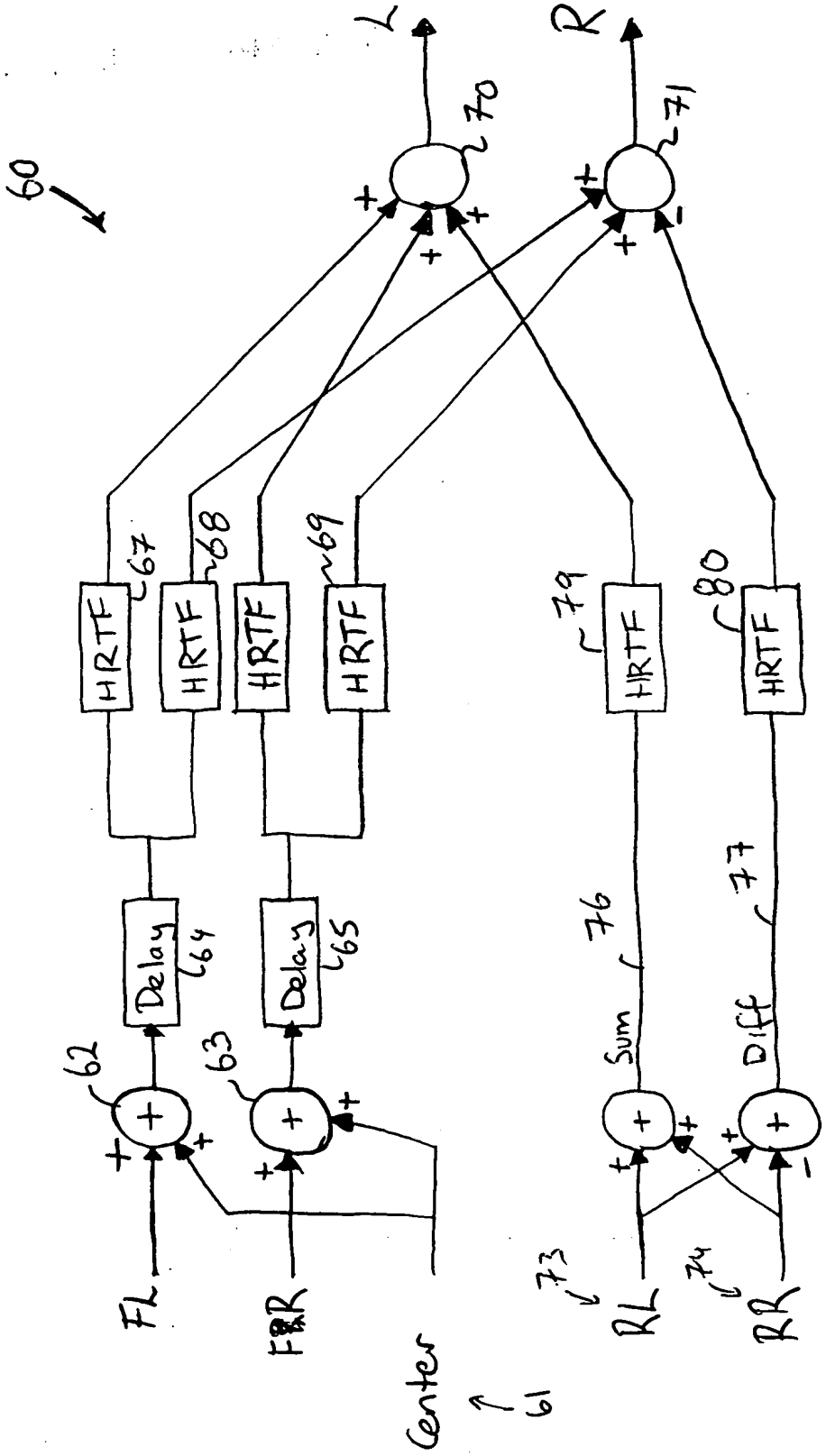
60
L
R
70
71
HRTF
67
68
69
79
80
Delay
64
65
62
63
Sum
76
Diff
77
FL
FR
Center
61
RL
RR
73
74

FIG. 6